Remote Sensing of Environment xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Multi-sensor detection of forest-cover change across 45 years in Mato Grosso, Brazil

Julie A. Fortin, Jeffrey A. Cardille*, Elijah Perez

Department of Natural Resource Sciences and McGill School of Environment, 21,111 Lakeshore Road, Ste. Anne de Bellevue, Québec H9X 3V9, CANADA

ARTICLE INFO

Edited by Marvin Bauer

Keywords:
Land cover
Change detection
Deforestation
Brazilian Amazon
Bayesian statistics
Roosevelt River
Mato Grosso

ABSTRACT

The ongoing march toward freely available, highly pre-processed satellite imagery has given both researchers and the public unprecedented access to a vast and varied data stream teeming with potential. Among many sources, the multi-decade Landsat archive is certainly the best known, but legacy and current data from other sensors is available as well through the USGS data portals: these include CBERS, ASTER, and more. Though the particular band combinations or non-global missions have made their integration into analyses more challenging, these data, in conjunction with the entire Landsat record, are available to contribute to multi-decade surveys of land-cover change.

With the goal of tracing forest change through time near the Roosevelt River in the state of Mato Grosso, Brazil, we used BULC and Google Earth Engine to fuse information from 13 space-borne imagers capturing 140 images spanning 45 years. With high accuracy, the resulting time series of classifications shows the timing and location of land-use/land-cover change—both deforestation and regrowth—at sub-annual time scales. Accuracy estimates showed that the synthesized BULC classification time series was better than nearly all of the single-day image classifications, covering the entire study area at sub-annual frequency while reducing the impact of clouds and most unwanted noise as it fused information derived from a wide array of imaging platforms. The time series improved and gradually sharpened as the density of observations increased in recent decades, when there were three or more clear, higher-resolution views of a pixel annually from any sensor combination. In addition to detailing the methodology and results of multi-source data fusion with the BULC approach, this study raises timely points about integrating information from early satellite data sources and from sensors with footprints smaller than Landsat's. There are decades of research deriving sensor-specific techniques for classifying land use and land cover from a single image in a variety of settings. The BULC approach leverages the many successes of single-sensor research and can be used as a straightforward, complementary tool for blending many good-quality mapped classifications from disparate sources into a coherent, high-quality time series.

1. Introduction

Deforestation plays an important role in the global carbon budget and in biodiversity loss, particularly in the tropics (Baccini et al., 2012; Barlow et al., 2016; Harris et al., 2012; Houghton, 1991, 2012; Whitmore et al., 1992). In response, international initiatives have been launched to ensure the sustainable management of forests worldwide, making the monitoring of forest cover increasingly important for researchers and decision makers alike. The development of remote sensing strategies to monitor forest cover across large areas has become a significant priority (De Sy et al., 2012; DeFries et al., 2007; Hansen and Loveland, 2012).

Using large data sets, researchers have made excellent progress in detecting changes in land use/land cover (LULC) in the new, open-data era (Wulder et al., 2018). The best-known existing algorithms typically use carefully calibrated imagery, most often Landsat-5, -7, and -8, to compare bands or build robust indices that are differenced or otherwise interpreted through time for stability and change (see especially the review by Zhu, 2017). Some use multiple sensors of the Landsat series (e.g., Zhu and Woodcock, 2014a; Zhu and Woodcock, 2014b) and some merge data of different resolutions (e.g., Hansen et al., 2008; Xin et al., 2013). Meanwhile, these algorithms have demonstrated their power at the global scale (notably Hansen et al., 2013), and hold excellent potential for detecting changes more subtle than wholesale land-cover change through abrupt disturbance, including abandonment (Yin et al., 2018), forest disturbance from low-density development (House and

* Corresponding author.
E-mail address: jeffrey.cardille@mcgill.ca (J.A. Cardille).

https://doi.org/10.1016/j.rse.2019.111266
Received 10 April 2018; Received in revised form 14 June 2019; Accepted 17 June 2019

Wynne, 2018) and forest structure change (Savage et al., 2018).

The Landsat archive is indisputably powerful, and could likely be made even more useful by integrating the vast amount of information from other, perhaps substantially different sensors. For algorithms that interpret raw imagery, sensors whose bands are fundamentally not equivalent to Landsat's optical bands (e.g., radar data from Sentinel-1) would be difficult to graft onto an index-based framework. Yet there is a still-increasing array of image types and derived products now available for interpretation and integration—for example, the National Agriculture Imagery Program (NAIP) photography (USDA FSA, 2018), agricultural censuses (Cardille et al., 2002), Hyperion (Pearlman et al., 2003), and LIDAR (Brennan and Webster, 2006; Wulder et al., 2012). In addition to algorithms that harmonize intercomparable image data for analysis, the growing breadth of sensors invites the development of algorithms that can track change and stability through time without regard to each platform's remote-sensing characteristics.

To address the challenge of fusing information from an arbitrary set of imagery sources, we have developed the BULC (Bayesian Updating of Land Cover) algorithm. BULC was first demonstrated in the context of tracking a fast-growing forest fire over a single summer season employing 11 Landsat-8 images (Cardille and Fortin, 2016). Despite using only one sensor in that study, BULC is equally capable of accepting classified imagery from a variety of sources. Because it operates on classified images rather than raw spectral values, example data sources might encompass optical imagery, radar-based or LIDAR-based classifications, existing LULC products like the US National Land-Cover Data Set (Homer et al., 2007; Vogelmann et al., 1993) and GlobCover (Bontemps et al., 2011), or at the most extreme, classified hand-drawn historical maps. Because there have been decades of work perfecting classification techniques for each of the sensors that are or once were in operation, analysts can benefit from past research by producing quality classifications from the perspective of a given sensor, and then synthesizing the multiple classifications in a time series produced by BULC. This study illustrates BULC's data fusion ability to track changes in forest cover in a nearly 12,000 km^2 study area over five decades, using the widest range of remote sensing platforms available to us during the period.

The Brazilian state of Mato Grosso, the Brazilian Legal Amazon, and the Amazon basin have been of substantial international concern due to the rapid rate of deforestation and other land-cover conversion in the past half-century (Fearnside, 2005; Malingreau et al., 2012; Saatchi et al., 1997; Skole and Tucker, 1993). Mato Grosso was the state within the Brazilian Amazon with the highest or second-highest deforestation area (km^2/yr) every year from 1988 to 2017, and has accumulated the second-greatest forest loss (km^2) since 1988 (total area of 1.43×10^5 km^2), or one third of the total forest loss in the Brazilian Legal Amazon (Câmara et al., 2006; Houghton et al., 2000; Instituto Nacional de Pesquisas Espaciais (INPE), 2017). Conversion began early in the satellite record: Cardille and Foley (2003) estimated a 4.7 Mha expansion of crop and pasture land cover in the state of Mato Grosso between 1980 and 1995, a 50% increase over the period. Subsequently, from 1995 to 2005, Mato Grosso contributed 33–43% of the annual deforestation increment in the Brazilian Amazon (Instituto Nacional de Pesquisas Espaciais (INPE), 2017).

Within Mato Grosso, the area surrounding the Roosevelt River is ideal for exploring the synthesis of decades of observations of forest change and stability across the entire satellite record. In the early 1900s, this region was one of the most remote on Earth, and was so named after the first complete navigation by non-Amazonian natives, in a group with varied trailblazing experience that included former U.S. President Theodore Roosevelt. Alongside Cândido Rondon (after whom nearby Brazilian state Rondônia was named), Roosevelt and his fellow explorers navigated the many rapids over several dangerous months in unworkably heavy canoes (Millard, 2005; Roosevelt Memorial Association Film Library, 1928; Roosevelt, 1914a, 1914b). With no signs of visible European-style settlement along the river's shore and no contact with the outside world, the trip nearly killed Roosevelt when a wound became infected while they were impossibly distant from recognizable landmarks or sources of medical help. Given that this river region was so remote and apparently unfarmed when Roosevelt's party struggled through in 1913–1914, what has it looked like in the satellite era, which began about a half-century after Roosevelt's journey? In this work we returned to that expanse of wilderness to explore the observable satellite record of its development history. Assembling a large number of relatively clear images from 13 different remote-sensing platforms – about three per year – we used BULC in Google Earth Engine to map forest, disturbed areas, and water at sub-annual intervals from 1972 to 2016. The resulting time series illustrates a region's LULC history across the entire satellite era, and outlines an approach that could be adopted widely to create high-quality LULC time series.

2. Methods

2.1. Study area

The study area is a rectangle of 96 km × 124 km (1.19×10^4 km^2), centered at 60.72° W 9.99° S along the Roosevelt River, in the state of Mato Grosso, Brazil (Fig. 1). At the beginning of the study period, the study area was almost entirely apparently undisturbed forest flanking the Roosevelt River and its tributaries, with an extremely small area that may have been already in agricultural use by 1972. We tracked three LULC types that were readily distinguishable in the study area. The first, “Forest”, was defined as lands dominated by woody vegetation with a percent cover > 60% and height exceeding two meters. This category encompassed the well-known IGBP forest categories 1–5 (Loveland and Belward, 1997). The second, “Water”, consisted of rivers and streams and corresponded to IGBP class 17, “Water bodies”. The third class, termed “Disturbed”, represented lands that, during the study period, were cleared for human use, including: croplands; forest clearing for pasture; regrowing lands that did not yet meet the definition of Forest; second or third clearings for agriculture; and roads. This corresponded to classes 6–10, 12, and 16 in the IGBP framework—the shrubland, grassland, savanna, cropland, and barren classes. As part of the classification phase described in a later section, we also identified a “No Data” category to encompass cloud pixels, shadow pixels and pixels with no image coverage.

2.2. Imagery

We assembled 140 images , Fig. 2) from 13 different sensors across five decades (1972–2016) for classification and use in BULC to understand change and stability in this region across the satellite record. The 115 Landsat 1–8 images are hosted at Google Earth Engine (Gorelick et al., 2017), and cover the entire study area. Four Sentinel images, also a part of the Earth Engine data catalog, were accessed for the area: the radar image from Sentinel-1 covered the full study area; three from Sentinel-2 were smaller tiles covering about 50% of the study area. From the USGS Earth Explorer, we accessed and uploaded to Earth Engine: seven images from the 20 m China-Brazil Earth Resource Satellite (CBERS) sensor (median coverage 81%), and 12 ASTER images (median coverage 26%). We also retrieved a true-color image taken by camera from a window of the US Space Shuttle and one taken from the International Space Station (Earth Science and Remote Sensing Unit. NASA Johnson Space Center, 2018).

The full image set spans 1972 to 2016 with variable temporal density (Fig. 2). As expected, useable observations were considerably more sporadic and noisier prior to the mid-1980s with the spatial and spectral improvements beginning with the advent of Landsat-5. It is important to note that because BULC blends classifications across multiple inputs, we could accept images that included clouds if some of the area was likely to produce a good LULC classification, as described in the next section. Where possible we identified three clear or

Fig. 1. Roosevelt River region and surrounding area, Mato Grosso, Brazil.

relatively clear images in each calendar year; some of the earliest Landsat years had only one or two available due to persistent cloud cover and only a single sensor returning imagery. Two years (1974, 1983) had no images or none clear enough to be used; this meant that across the 45-year period we were able to update an estimate of LULC in 43 distinct years. When data from the entire period were assembled, the years 2004–2006 were notable for the density of imagery: Landsat-5, CBERS, and Landsat-7 were each providing imagery in that period. Images were registered using Earth Engine or ArcMap as needed and classified for use in BULC as described below.

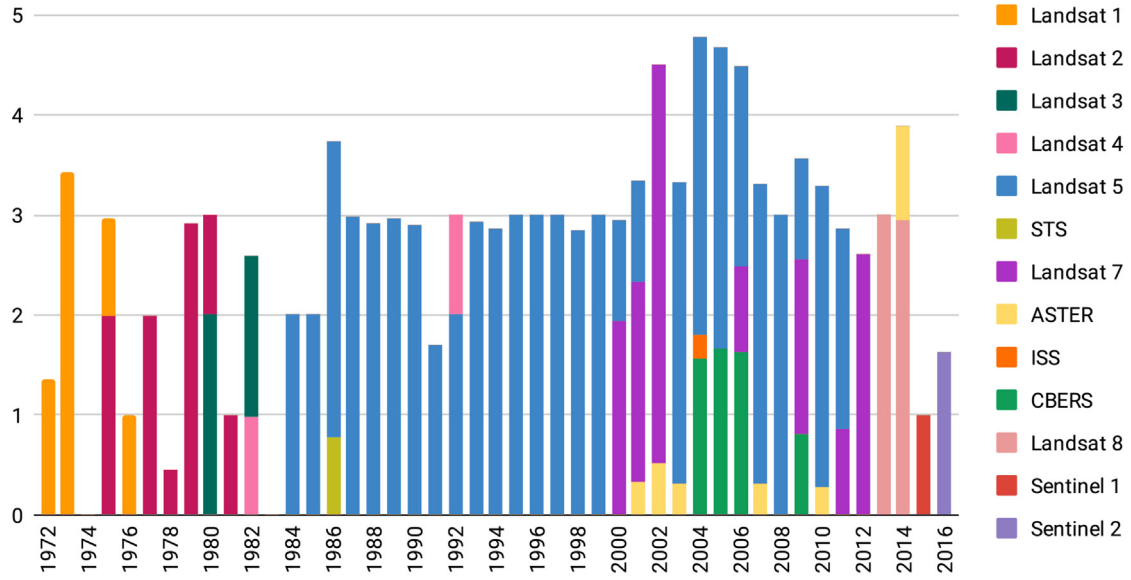
2.3. Event creation

As described in Cardille and Fortin (2016), BULC processes a series of LULC classifications (termed "Events"), each of which can be produced using any method suitable to interpreting images from the given sensor. Here, we used the Earth Engine "Explorer" interface to create an Event from each image using classification and regression trees (CART) (Breiman et al., 1984), one of the built-in classification techniques in Earth Engine.

The Explorer interface enabled the delineation of small regions and points of uniform LULC to train classifiers. For each of the categories, we created small sets of training points and polygons through on-screen digitization and labeling of homogenous areas. Where possible, we used the same set of training data in creating each Event, similar to other efforts to identify "pseudo-invariant" sites to facilitate high-quality mapping of images from multiple dates and platforms with stable training sets (Fortier et al., 2011). We labeled points when their proper class changed (for example, from Forest to Disturbed) during the study period, and occasionally added points and polygons to aid the classification of a given image for an Event. For imagery such as ASTER whose footprint covered only a small part of the area, we developed a distinct set of training points sufficient to produce a reasonable classification in Earth Engine with CART for an image of that footprint on that date. As the final step of Event production, a focal mode filter with a radius of 2 pixels was applied to each classified map. Because BULC is designed to work on a large number of Events that each are of moderate but not exceptional quality, we performed classifications relatively quickly without the usual goal of making each classification nearly perfect. Taken as a whole, the set of Events presented a view of the study area through time that contained sporadic sensor errors (Fig. 3a), partial-coverage imagery (Fig. 3b), systematic stripes (Fig. 3c), some properly removed clouds, some incorrectly classified clouds, and good-quality classifications that covered the full area (see also Appendix 2).

2.4. BULC algorithm

As detailed in Cardille and Fortin (2016), the BULC algorithm maintains a running estimate of the probability of each of several tracked classes, as initialized to a plausible starting point and observed



Sensor	Mean Spatial Resolution (m)	Number of Images	Mean Coverage of Study Area (%)
ASTER	15	12	23
CBERS	20	7	81
ISS	27.31	1	24
Landsat 1	70	8	85
Landsat 2	76	10	93
Landsat 3	80	4	90
Landsat 4	45	2	100
Landsat 5	30	69	99
Landsat 7	30	15	93
Landsat 8	30	6	99
Sentinel 1	5	1	100
Sentinel 2	20	3	54
STS	541	1	78

Fig. 2. Above: temporal distribution of data in the Roosevelt River study area, 1972–2016. A value of 1 means that a typical study area pixel was, on average, imaged once in a given year. Below: spatial resolution, image count, and average coverage of the study area for each sensor. Original MSS resolution was 79 m × 57 m; some images were resampled to 80 m and some to 60 m, accounting for the averaging shown. Landsat-4 included images from both its MSS and TM scanner.

through time in an ordered series of Events. BULC ingests already-classified images from any data source, meaning that there was no need for special treatment for Events derived from, for example, Sentinel-1, Landsat-4, or ASTER. Not all images affect the probability estimates evenly: higher-accuracy classifications, and more specifically, higher-accuracy classes within those maps, have larger impacts on probability estimates.

To create its time series, BULC uses the agreement between Events as an indication of the reliability of each new classification, to estimate the LULC of the entire study region based on the evidence from all Events up to that point. To understand how Events are processed in BULC, consider a given Event in which Water was exceptionally well classified with very little omission or commission error, and that Forest and Disturbed were well classified but occasionally confused with each other. As detailed in Cardille and Fortin (2016), BULC creates conditional probabilities at each time step from an overlay of each given Event with the previous Event in the set. The conditional probabilities of each class are computed using the reproducibility of each (1 – omission error) from the previous Event to the current one. The effect is that classes that are well repeated from Event to Event are considered as strong evidence by BULC's Bayesian calculations. In the example Event, for any pixels classified as Water, the probability of Water increases considerably because that class is highly reproducible between the given Event and the previous one. With the probability of Water increasing, the probability of Forest and Disturbed necessarily decrease, since Water was little confused with either. For a pixel classified in the same Event as Forest, the effect is similar but different in important ways. For Forest pixels, the probability of Forest would increase by an amount related to the reproducibility of the Forest category, which in this example is lower than that for the reproducibility of Water. Because Forest was occasionally confused with Disturbed but very rarely with Water in this example, the probability of Water would decrease

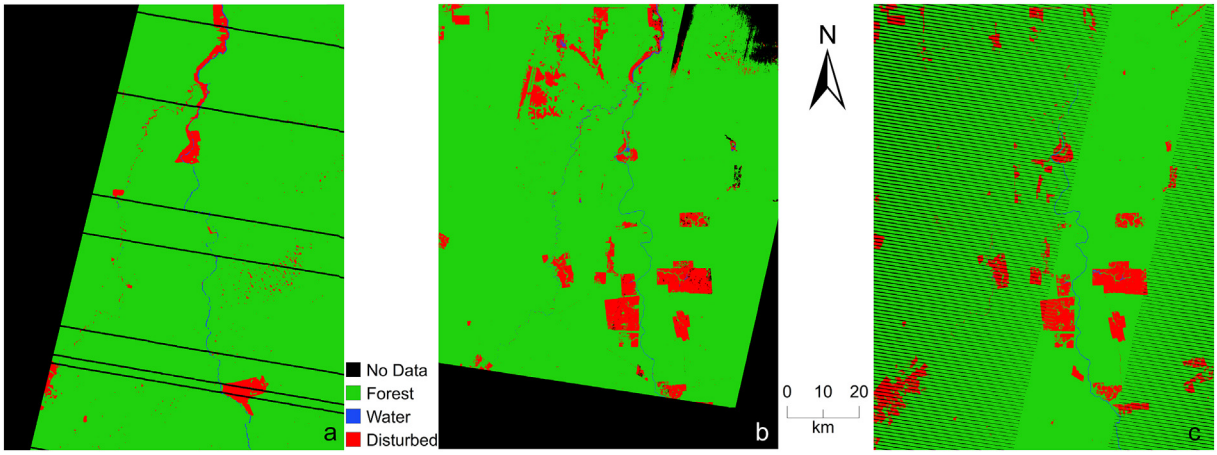


Fig. 3. a, b, c: Classified events for July 31, 1982 (Landsat-3, 80 m), August 18, 2006 (CBERS, 20 m), and June 15, 2012 (Landsat-7, 30 m) respectively.

substantially, while the probability of Disturbed would change more modestly. In practice, the probability vector of {Forest, Water, Disturbed} might change from {0.65, 0.04, 0.31} to {0.72, 0.02, 0.26} in a pixel where the Event value was Forest. The specific change of values is driven by the conditional probabilities and also related to the value at the beginning of the iteration, since values asymptote near 1 and must add to 1 at the end of each time step. At a given time step, only those pixels with new evidence (i.e., that are not No Data in the given Event) have their probabilities updated: probabilities where a given Event is No Data are left unchanged for that iteration. Additional detail and examples are given in Cardille and Fortin (2016).

Three parameters drive the BULC tuning: First, a parameter initializes the probabilities for the first image (here set to 0.8 for the class of the first Landsat classification and 0.1 for the other two). Second, a 'leveling' value that dampens extreme values of the conditional probabilities. This was set to 0.4, which modestly inhibits extreme changes that otherwise would have pixels changing their estimated class in response to classification noise. Third, because Bayes' formula can get stuck if a probability becomes 0, we ensure that each of the three categories has a small nonzero probability at the end of each iteration (using the parameter value 0.95). Together, the three parameters served to govern the process by trusting the first classification and incorporating new Events as good but not perfect evidence. These settings were able to change a pixel's estimated class after between two and four Events, depending on the agreement between classifications and a pixel's history.

A considerable strength of BULC is that it automatically manages a full set of LULC probabilities in the context of new data- that is, its function is to decide the specific amounts of increase and decrease in probabilities of each LULC type in light of evidence from a new Event. In this study, LULC for the beginning of the time series was initialized to the first view of the full region in the satellite era: July 29, 1972, just 6 days after Landsat-1's landmark launch. Data from sensors launched later were brought into BULC as they became available, in chronological order for a total of 139 iterations. We tasked Earth Engine with producing its outputs at 20 m spatial resolution for estimation of LULC proportions, changes through time, and accuracy assessment.

2.5. Accuracy assessment

Accuracy measures of the BULC classifications and Events were computed on a reference data set of 692 points. Candidate points were selected via a probability sample (Stehman, 2009) from three classified images in the following way. For the main body of points, we stratified randomly from the Event classification of August 29, 2013, identifying 225 Forest points, 200 Water points, and 117 Disturbed points. Because only a small portion of the Disturbed area in 2013 was already disturbed in the 1980s, and because only a small portion of the Disturbed area in the 1980s had been disturbed in the early 1970s, we added additional reference points to capture that earlier human activity. This was done by randomly selecting 75 points in the Disturbed category for the Event of August 1, 1980, and 75 more randomly selected points from the Disturbed stratum from the August 29, 1973 Event.

Each of the 692 points was estimated at each of the 140 dates in the following way. Because there were no separately gathered ground observations available in any of these points at any of the time steps, we estimated the reference LULC class (Forest, Water, Disturbed) based on visual inspection of the corresponding unclassified imagery. The evaluator interpreted the LULC at 10 m in the cardinal directions around each identified sample point, making a sample unit that simulated a pixel size of 20 m. In the rare case that a simulated pixel spanned two LULC categories, the evaluator estimated the majority LULC as the reference value. In practice this had little effect on the accuracy estimates since few points were near LULC edges given the sizes of features, particularly of the disturbed areas and forests. Labeling of the points at each time step was done without reference to the Events or resulting BULC classifications. Due to cloud cover or missing data within some images (as in Landsat-7 SLC-off imagery), there could be some ground points within a given image's footprint that were empty or otherwise impossible to discern by inspecting the image. Because LULC was not ephemeral at the sub-annual time scale (e.g., did not change from Forest to Disturbed and back again between two viewings), we applied the same label for an unviewable point on a given date if the LULC appeared to have been unchanged between the preceding and following images. This permitted a full set of reference points to be used to evaluate each BULC classification at each time step, for a total of about 98,000 observations.

We calculated two formulations of accuracy: (1) the accuracy values outlined in Congalton (1991) and (2) the unbiased estimates as revisited since by others (Olofsson et al., 2014; Stehman, 2014; Strahler et al., 2006). At each iteration, we assessed each Event and each BULC classification with the maximum amount of assessment data available for that footprint. When an Event did not cover the study area, only the accuracy data from within that Event's footprint could be compared to the Event LULC estimates. Because the BULC algorithm produced an estimate in each study area pixel at each time step, however, the full set of accuracy assessment data was always used for the BULC time series. This had the intentional effect of raising the accuracy challenge for the BULC time series—it had the responsibility of high-quality mapping at every time step over the entire study area, even in areas that may have been undergoing LULC change but whose status could not be seen but only inferred at that time step. Though our protocol could not entirely eliminate the risk factors encountered when assessing LULC

classification in the tropics noted by Powell et al. (2004), we worked to minimize avoidable obstacles to realistic assessment of the Events and BULC classifications. These efforts included using clear categories, well-registered images, and LULC that does not change between a map's nominal date and the date of reference data.

2.6. Extracting LULC sequences

BULC produces a time series of LULC categories as its output, but does not assign meaning to the sequences of LULC categories seen throughout the study area. To detect dynamics of deforestation, regrowth, and reconversion, we developed a sequence-matching tool in Earth Engine to identify the location and timing of these three noteworthy LULC change types. The sequence matcher inspected the BULC classification value of each pixel at the end of each year. First, to identify a conversion of Forest to Disturbed, we identified sequences in which at least three years of Forest were followed by three years in the Disturbed category as assessed in the year-end BULC classes. The second type of LULC change tracked in the annual BULC classifications was the reversion of Disturbed back to Forest. This was based on a sequence of three years of Disturbed followed by three years of Forest. Because the appearance of regrowing Forest in images often appeared for a time to exist in an indeterminate state between Forest and Disturbed, some alternate sequences of reforestation were observed and retrieved as described in Appendix 3. The final type of LULC change was a reconversion to the Disturbed category, as a second (or, much more rarely, a third) clearing after reforestation. To find any reconversions, we identified pixels in which the deforestation sequence (three years Forest, three years Disturbed) was again observed after an initial, earlier clearing and sustained regrowth. We developed the set of reforestation sequences by closely inspecting fields that, to the eye, were undergoing the same LULC history, but in which a few pixels had similar but slightly non-standard LULC histories in Events and BULC. After a sequence was detected, the year of the change (such as from Forest to Disturbed) was recorded.

3. Results

3.1. BULC time series for the Roosevelt River region

The time series of Events, before introducing them into BULC, was not a realistic portrayal of true LULC change in the Roosevelt River region. Because of the strengths and weaknesses of different sensors, the amount of cloud and haze, missing data, and classification error, individual Events of the time series differed considerably from each other and could not be simply superimposed and analyzed as a meaningful, realistic time series (Appendix 2). A given pixel might be classified in the Events as Forest on one day, Water the next, then Disturbed, then Forest again. When considered as a series, the individual Event classifications disagreed on important, basic characteristics like the proportion of Forest cover—in two consecutive 1991 Events, for example, one Event estimated 94.4% Forest and the next, from a few weeks later, estimated 95.9%. Even more importantly for mapping a coherent time series, the set of Events disagreed substantially among themselves on the locations and patterns of the LULC classes within the study area.

BULC interpreted the information contained in the Events to produce a continuous, coherent LULC time series for 140 imagery dates between 1972 and 2016 (Appendix 4). The series assessed, for each time step (e.g., for the dates of Fig. 4), its best estimate in every pixel, even those that were not imaged for a given time step (corresponding to the No Data areas in Fig. 3). With updates on average several times each year, the series showed the varieties of LULC change dynamics over the entire satellite era. These included new deforestation between 1982 and 2006 in the southwestern part of the study area; abandonment and reforestation that split a large field in the southeast into two; small, isolated clearings amid large, contiguous ones in the center of the study area; and deforestation, regrowth, and subsequent reclearing (e.g., in the extreme north center of the study area).

The BULC time series estimated that the proportion of Forest in this study area fell from nearly 100% in 1972 to around 93% in 2016, a net forest loss of 795 km^2 (Fig. 5). The unbiased assessment of map areas (Stehman, 2014), which uses the characteristics of the accuracy assessment points to account for sampling bias and the size of sampled strata, estimated that there was even less forest in the region at the end of the time period: 90.6% in 2016, for a net loss of 1068 km^2. Although the downward trend in forest cover was very roughly linear across the time series, there were several periods at which net forest loss occurred at higher rates: notably, between 1979 and 1982, adjacent to the Roosevelt River in the north of the study area; between 1996 and 2000 in the center of the study area; and in 2004–2005 throughout the region, after which deforestation continued but slowed considerably, at least until this study's 2016 end date (see an animation of the time series in Appendix 4).

3.2. BULC and event accuracy

The Overall Accuracy of the entire BULC time series averaged 98% in the unbiased calculation and 90% in the standard formulation (Fig. 6, Table 1), well above the recommended threshold of 85% (Foody, 2002; Thomlinson, 1999). Average per-class Producer's and User's Accuracy values were similarly high (Table 1), though with some aspects that merit further discussion below.

The Producer's Accuracy for the Water class (Fig. 6, second panel) illustrates several important aspects of the BULC time series. Stretches of several rivers in the Roosevelt region were too narrow to be classified reliably as Water in MSS imagery: in the early part of the satellite record when MSS images provided the only satellite data, narrow rivers were typically classified as their surrounding Forest class. This limitation was reflected in the BULC time series (Fig. 4, left), in which large stretches of narrow rivers were classified as Forest until the arrival of TM images (Fig. 4, middle, particularly in the southwest part of the study area). These resolution-driven constraints can also be easily seen in Appendix 4.

When the accuracy of the BULC Water class was estimated with both the standard Congalton (1991) and the unbiased Stehman (2014) formulas, different opinions emerged (Fig. 6, second panel). In a series built from MSS Events that had quite high omission error, the Congalton values record that 60%–75% of the true water pixels were properly classified as Water during the MSS era; this matches what is seen at finer viewing magnifications and the informal observation that one river was wide enough to be mapped with MSS, while the other was mostly not. Soon after TM imagery was introduced in the mid-1980s, BULC's ability to discern these rivers quickly jumped to around 90%, where it remained stable. In contrast, the unbiased accuracy estimates were susceptible to large, unrealistic jumps and plateauing (Fig. 6, second panel). This is because sample points of the very small Water class were sometimes near to the very large Forest class, which made the unbiased accuracy estimates unstable based on the behaviour of a few accuracy assessment points. In our judgement, the Congalton calculations provide a more realistic and better-behaved estimate of omission error for the Water class.

For the Disturbed class, the BULC time series had low omission error (Fig. 6, third panel) for most of the time series—from about 1980 until the present. Early in the satellite era, however, the Producer's Accuracy began low–not because of poorly classified imagery, but because of updating delays due to infrequent imagery, during a period when the amount of Disturbed area happened to be rapidly increasing. BULC had been parameterized to confirm new disturbances only after several images had been collected, which equated at that time to a year or more. In the study area, the most apparent omission in the 1970s was located in the southeastern part of the study area, in which a very large forest clearing in mid-1976 (as evidenced by eye on the Landsat image

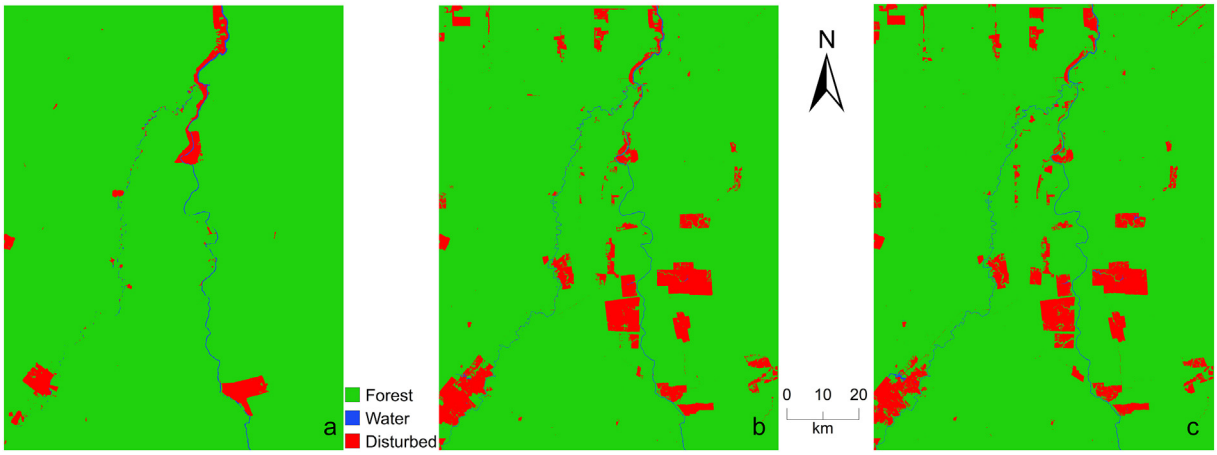


Fig. 4. a, b, c: BULC time series excerpted for July 31, 1982, August 18, 2006, and June 15, 2012, corresponding to the same panels in Fig. 3.

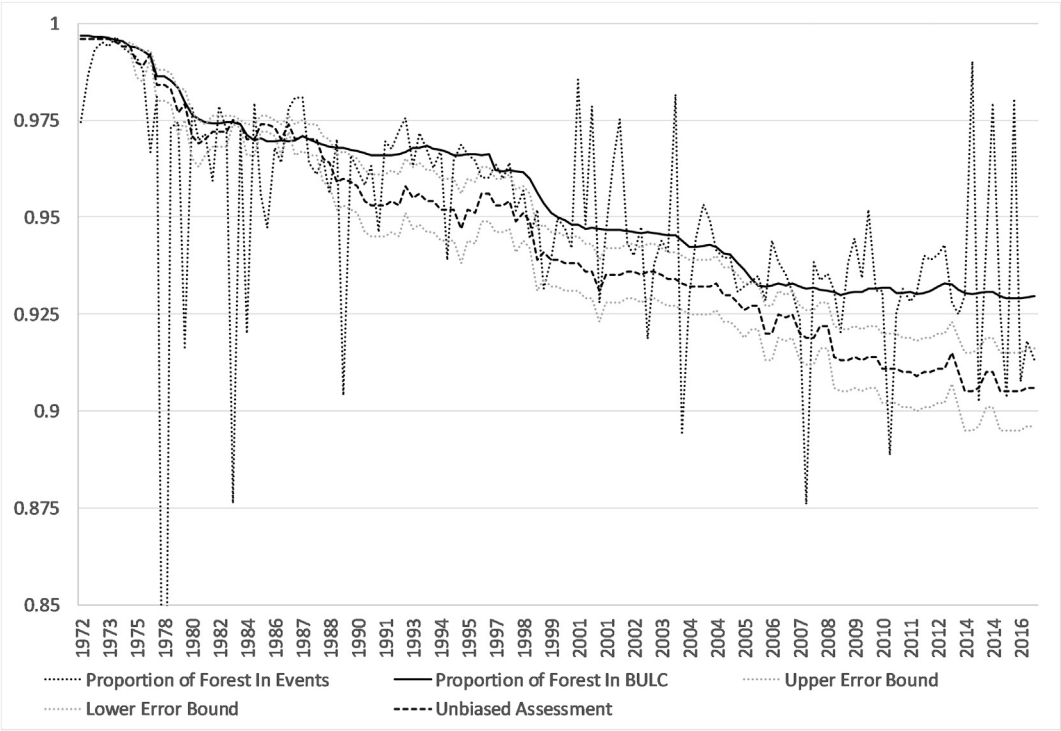


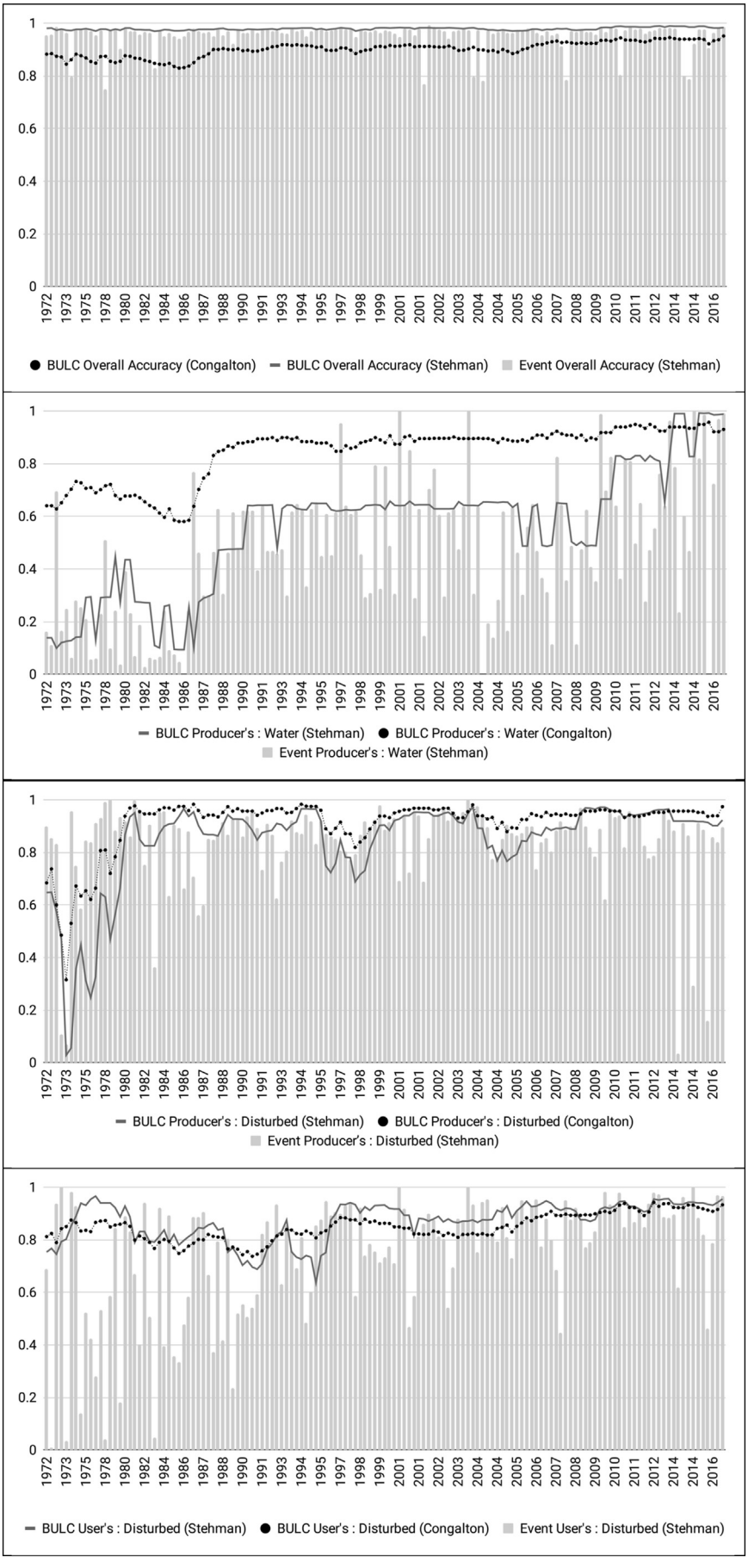
Fig. 5. Forest cover proportion in the study area from 1972 to 2016 time as estimated by individual Events (dotted dark gray lines) and the corresponding BULC classifications (solid line). The unbiased estimate of forest cover proportion (dashed line) with calculated confidence intervals (dotted light gray lines) indicates even less forest than that seen on the BULC classification maps. Note: because very small Events (such as those made with ASTER) may cover only part of the study area, some of the volatility of this proportion in Events in later years was due to smaller footprint: Events covering a smaller area were more likely to have an unrepresentative proportion of forest.

and the Event, as visible in Appendix 4) was not confirmed until mid-1977. This time lag was manifested as a low Producer's Accuracy, which resolved itself only when new relatively clear imagery later arrived quickly enough for BULC to catch up to the then-current LULC pattern of Disturbed area.

The User's Accuracy of the Disturbed class (Fig. 6, fourth panel) is informative for several reasons. First, BULC is able to synthesize LULC maps of Disturbed areas from very low-quality inputs, especially for the period prior to 1991 (compare the Event bars of Fig. 6 with the accuracy estimates). In our investigation of the accuracy assessment results, most of the commission confusion did not come from entirely undisturbed areas that were misclassified, but rather was principally due to evaluator uncertainty about when abandoned or fallow agriculture land had regrown sufficiently to be properly called Forest. With a substantial proportion of the Disturbed area in some form of recovery throughout the process, this uncertainty in labeling the reference data drove what was marked as commission error. There were a few places marked as Disturbed that were never farmed (rock faces, permanent sandbars), but these were a diminishingly small part of the signal as agricultural conversion continued.

3.3. LULC trends through time in the Roosevelt region

The BULC time series indicates that between 1972 and 2016 a total of 980.46 km^2 of the study area was converted from Forest and kept in the Disturbed category for at least three years, amounting to about 8.3% of the study area. Abandonment of disturbed land for a long enough period to revert to the Forest class was seen in 27% of lands



(caption on next page)

Fig. 6. Accuracy of the BULC and event classifications through time, for each of 139 events between 1972 and 2016. First panel: Overall accuracy for the BULC time series and Events. Second panel: Producer's Accuracy for the Water class. Third panel: Producer's Accuracy for the Disturbed class. Fourth panel: User's Accuracy for the Disturbed class. The complete set of accuracy graphs is given in Appendix 5.

Table 1
Accuracy values for the 1972–2016 time series averaged across the 139 BULC classifications. Two values are shown for each accuracy component: first the Congalton (1991) values; second, in parentheses, the unbiased adjustments made following Stehman (2014). As described in the text each component captures a different aspect of accuracy.

	Omission (producer)	Commission (user)
Forest	0.92 (0.99)	0.91 (0.98)
Water	0.84 (0.56)	0.95 (0.94)
Disturbed	0.92 (0.78)	0.88 (0.82)
Overall	0.91 (0.97)	

cleared once; about 22% of those reforested lands were later cleared again (Fig. 7).

The reforestation proportion was comparable to but somewhat lower than that estimated in other studies across much larger areas (e.g., Cardille and Foley, 2003; Houghton et al., 2000). The discrepancy may illustrate geographic differences, but more likely reflects the relatively restrictive definition of reforestation in this study—that is, land recovered enough to be classified again as Forest for a sustained period. The estimated amount of reforestation is not a function of BULC per se; rather, it is a reflection of several factors, including the definitions of deforestation in the post-processing sequence matcher, the LULC categories chosen for classifying Events and the sensitivity of the classification algorithms and training to those classes. Analysts wanting to do a more specific evaluation of regrowth and subsequent clearings could produce a new series of Events that took specific care to identify LULCs of interest, e.g., recent abandonment, second-growth forest, pasture, and others. Those Events could be combined in BULC in the same way described here, and interpreted accordingly.

The analysis of the full satellite record can be used to see the beginnings of conversion for modern agriculture in the Roosevelt River area. The complete time series of BULC classifications shows the development of the region from a nearly unfarmed area to one undergoing steady conversion to an agricultural landscape. This is consistent with the surrounding region, notably even more extensive conversions in Rondônia to the west and Mato Grosso to the east. The image of change complements and extends the record of Hansen et al. (2013), who found forest loss in the same parts of the study area in recent decades. The assembled multi-decade LULC history (Fig. 8, Fig. 4, Appendix 4) shows that conversion for agricultural uses since 2000 is broadly consistent with its longer history: the region has been developing continually since the 1970s.

4. Discussion

The synthesis of LULC observations from multiple sensors providing data of variable density, resolution, and accuracy raises several important points for discussion. These include: the strengths and weaknesses when considering noisy data from early in the time series when images were rare; factors influencing the quality of the BULC time series; interpreting the effects of changing spatial resolution through time and judging the value of images with small footprints; and limitations to the BULC algorithm in this setting.

4.1. Data quality and quantity

When considering how best to incorporate information from the earliest years of Landsat, we found a tension between the desire to include as many images as possible from the early, data-poor part of the time series and the instinct to avoid noisy classifications. Because of familiar limitations of spectral and spatial resolution in MSS imagery, early satellite images created poorer classifications than more recent ones, even into these few relatively distinct categories. On the other hand, a given image from Landsat 1–4 was a rare early look at the region and had a high potential, even given its likely flaws, to inform our understanding of LULC from that time. In effect, earlier observations were more valuable than more recent ones because of their rarity, and users might want to expend more effort to create Events from them. Ultimately, some users might be willing to experiment with including substantially cloudier imagery at the beginning of the sequence, or to

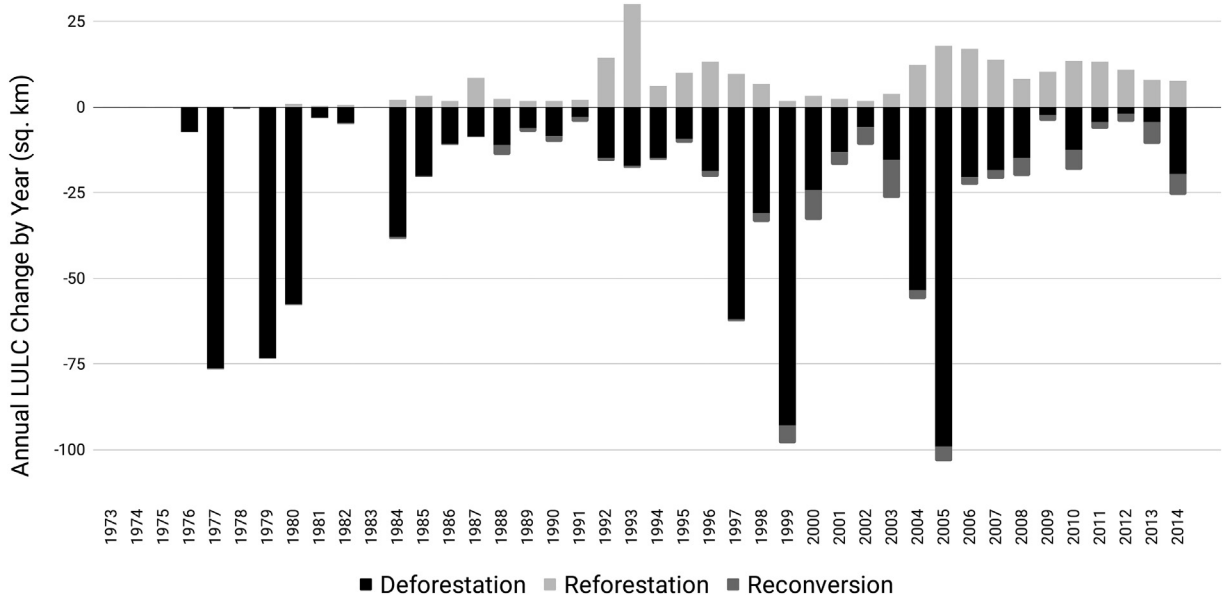


Fig. 7. Annual deforestation, reforestation, and reconversion (in km^2) in the Mato Grosso study area, 1975–2014. Deforestation and reconversion are shown as negative, reforestation as positive.

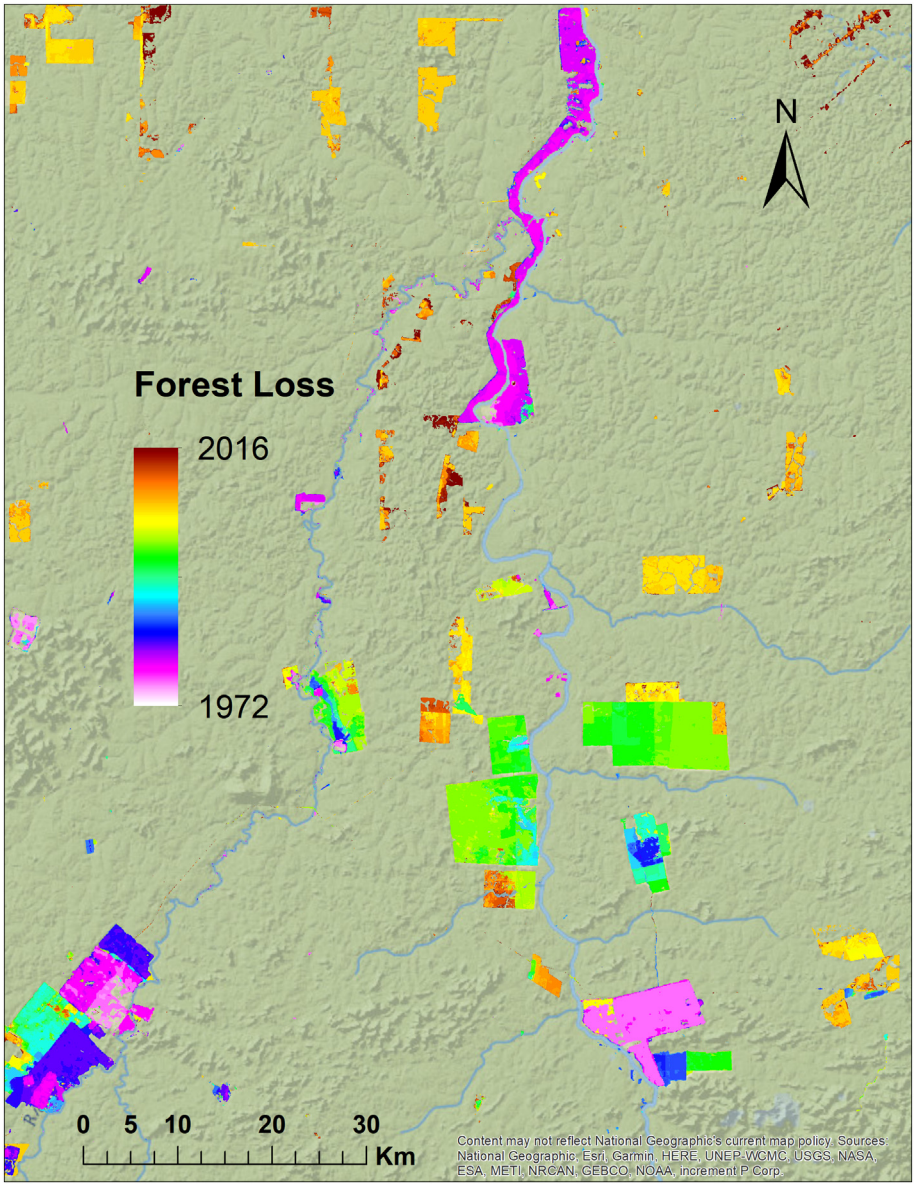


Fig. 8. Date of first clearing in Roosevelt River study area as detected in the BULC time series.

include more than a few clear images per year where possible.

Data frequency can affect BULC's detection and estimated timing of LULC changes. BULC is not limited to any maximum number of images per year or number of years for its functioning. In a situation where more images per year were available, this would give the ability for LULC change in those years to be evaluated more frequently. This would afford more opportunities to confirm either change or stability in LULC classes. We found that three (or more) images per year, as seen here, allows most changes to be identified within a year of when they occurred. Meanwhile, updates can continue indefinitely into the future as new evidence arrives. Where data is substantially less frequent, however, such as early in the satellite era or in extremely cloudy locations, timely updating is inevitably difficult. For persistent phenomena, limited data availability can greatly affect accurate timing of LULC. That is, a deforested parcel may eventually be detected and confirmed, even if the exact timing is not well known. For ephemeral changes, such as floods, sparse imagery may mean missing an LULC change altogether.

One of the intended eventual audiences of BULC is users with access to a platform like Earth Engine but without experience performing atmospheric corrections, discerning haze in imagery, devising radiometric corrections across sensors, and other significant training in remote sensing. Here, despite the quality of the BULC classification operating across multiple sensors, the time series did retain some transient noise. If desired, much of this noise could be removed with more meticulous preparation of Events, particularly through greater attention to detail in cloud and cloud-shadow masking at the stage of Event creation. Given BULC's success in producing a highly credible time series from many inputs that had severe limitations (Appendix 4), BULC may be able to make the production of long, accurate time series feasible for users of moderate remote-sensing experience.

BULC was parameterized to be relatively conservative in its tendency to accept new information, which had several effects. This parameterization meant that a possible LULC change needed to be confirmed several times, depending on the strength of the evidence contained in the Event classifications, before the probabilities flipped from one category to another. In the later part of the time series, when observations were plentiful, this conservative parameterization meant that a valid change (say, to Disturbed) was typically confirmed within the correct year. Earlier in the sequence, however, forest clearings were

similarly confirmed after a few images, but the rarity of images at that time in this area meant that this could be several years later. This had the effect of lowering the Producer's Accuracy of the Disturbed class early in the initial years, since at a given moment in the early part of the sequence, there were areas that had been converted but not yet imaged often enough to confirm changes. It would be possible, though we did not do it here, to parameterize BULC to treat evidence from certain images more (or less) conservatively than their conditional probabilities would indicate. Depending on a study's goals, a user might wish to weigh early images less strongly (because they are more likely to contain registration or classification errors) or more strongly, because a given observation of, say 75% confidence may be considered much more valuable than no observation at all, in years of rare imagery and potentially rapid change.

4.2. Changing minimum mapping unit through time

Although the images in the BULC time series all were exported with 20 m resolution, the features visible in the series changed as inputs sharpened over the 45-year series. Just as the size of the smallest features visible on different sensors has varied across the long satellite record, the BULC time series could resolve features of different sizes throughout the study period. At any point in the BULC time series, the minimum mapping unit is a product of the interplay between the spatial resolution of recent Events, the majority filter applied during Event creation, and the agreement between Events. Using the rule of thumb of an MMU being four pixels in an image and considering that the filtering of each Event removes single pixels, we estimate that the MMU during the MSS era was about 10 ha. For most of the TM era, the BULC output sharpened to an MMU closer to 1 ha. Near the end of the time series, in the period closer to the present as platforms have proliferated, estimating the MMU is more complex. The smallest resolvable unit is a function not only of the minimum input resolution provided, but also its frequency in the series—when finer-resolution Sentinel imagery was used frequently at the end of the period, the BULC time series took on those spatial characteristics and was able to resolve new features- for example, forest islands in disturbed areas and small reservoirs. But if Sentinel imagery were only available in one part of the study area, that sector would become sharper and the rest would retain the information visible on a 30 m image. (In contrast, if only MODIS imagery were provided to BULC, the output would take on the characteristics of MODIS data.) In today's potential mix of resolutions from Sentinel-1, Sentinel-2, Landsat, and others, the MMU of a given date could vary somewhat across the region depending on the mix of recent images over parts of the study area. From one perspective, that comports with a philosophy of using all available imagery to make the best map at every location. Variability in the MMU across space and time might not be appropriate for all applications; users who prefer a stricter control over the spatial characteristics of the BULC time series can either (1) resample all Events to a common, coarse spatial resolution; or (2) coarsen the BULC output to a common, coarse spatial resolution.

4.3. Initial conditions, event accuracy, and bias

Because BULC builds its time series by considering the accuracy of an Event versus the previous one, it might seem that the quality of the very first image would be paramount and drive the success of the entire BULC run. Somewhat surprisingly, this is not the case in practice. Because BULC's primary function is to change the state of pixels given new evidence, a poor initial map is quickly overwritten given a few Events that present consistent LULC evidence. In testing, we found that no trace of even an entirely random starting map was evident after about 6 images. That said, it was fortuitous to begin the sequence with the very first Landsat image, which was quite clear and created an excellent starting condition.

BULC follows the evidence from Events to characterize Earth's surface in a user's desired categories. Systematic errors, such as repeated misclassification of an area, in a set of Events cannot be discerned by BULC and would be interpreted as whatever is indicated by the Events. Given BULC's design, biases need to be monitored at the stage of first creating the Events. There is no need, from BULC's perspective, for Events to use relatively broad categories as seen in this study. Event creators might decide to invest the effort to create very fine-grained classes- consider forest age classes as an example. If analysts can reliably delineate a set of classes across images (granting occasional errors), BULC can work even very similar categories to fuse the classifications into a time series.

BULC is best suited for phenomena that have signals that persist—that is, that lasts longer than a few images. Where data is infrequent relative to the mapped phenomena, there is a risk of missing phenomena of interest. For example, floods may be of interest but if imagery is only available every few weeks, BULC is not the right tool if the flood only is visible in a single image. In this study, there was a concern that a short period of reforestation might be missed if it only occurred for a few years before being reconverted for human use. In this small area, we could confirm that was not a widespread problem. But in a larger area in which imagery was sparse and cloudy, transient states should be interpreted with care.

We have directly graphed and compared the accuracy of Events versus the BULC time series here, but that strategy has important potential limitations. Because many Event images had footprints of different sizes, each one that was smaller than the full study area intersected with a different subset of the reference data. We gave BULC the harder challenge of correctly representing the entire study area at all time points, comparing it against all available reference data at each time step. In this work, there were enough points to express the quality of each of the Events fairly. More generally, however, as BULC fuses data of many different qualities and resolutions, we are uncertain how best to weigh the contribution of individual images in time series classifications built from data of unique characteristics. What is a given ASTER image's value to the process, for example, in sharpening the classification to sub-30 m resolution if it covers only part of the study area and is of moderate accuracy? In a time series built over an enormous area, the question may be more pressing than it was here in a study where the image list could be built by hand. It may soon become important to be able to automate the weighing of several important criteria (footprint, resolution, accuracy) to judge whether to include an image in a time series. A strategy for formally considering Events and each one's effect on BULC classifications through time, accounting for changing footprints and resolutions, remains to be developed.

4.4. Accuracy in long time series

Accuracy assessment values were useful for a broad understanding of the utility of the Events and the BULC time series, but they were inherently limited in ways that, to us, seriously constrained their interpretation for comparisons or their use in isolation from viewing the Event and BULC time series. For example, the Event from August 18, 2006 (Fig. 3b) was evaluated as having a 95% Overall Accuracy, very near to the 97.5% result for the BULC time series on that date (Fig. 4b). Given these high and similar values one might conclude, especially if one could not view the images of the classifications, that the Event and the BULC item for those days were equally useful and likely quite similar. Yet the Event, despite its high score, has severe flaws that would limit its use as a representation of LULC across the study area at that time, except with additional context that could potentially be found by incorporating data from other images from the same period. These include apparent substantial commission errors in the Disturbed category, as well as no data at all for 19% of the study area.

Surprisingly, we found that the unbiased per-class assessments were unstable across the time series for the small Water and Disturbed classes. This was inherent in the methods of calculation of the unbiased

assessments (Stehman, 2014): reference pixels from these small classes that were misclassified as Forest had enormous potential effects on per-class estimates despite the underlying BULC maps being nearly indistinguishable (Fig. 6, Appendix 4). Except for these abrupt dips and jumps, the estimates of Congalton (1991) are similar in magnitude to the unbiased estimates in the Water and Disturbed classes, and we advise looking to the Congalton estimates to understand the stability of the BULC time series in those classes.

Sharpened resolutions over the satellite era tended to improve accuracy in the BULC time series, though not always. In the earliest part of the satellite record, the Water class revealed a demonstrable effect of improving spatial resolution from 80 m to 30 m. When compared to the 80 m MSS resolution, the Roosevelt River and the Rio Branco to its west were narrower than could be reliably resolved in our classifications. Given the MSS obstacles of coarse spatial resolution, low radiometric resolution, and residual registration uncertainty, the Producer's Accuracy for Water remained at levels we expected, respectable but relatively low (~60% in the stable Congalton estimates) for much of the early satellite period until Landsat-5 was launched. From that point forward, both User's and Producer's Accuracies were quite high as the borders of the rivers became increasingly well-defined through time once sensors with higher spatial resolution were introduced. Recent improvements in spatial resolution through time were only somewhat reflected in the standard accuracy values. The effective spatial resolution of the BULC time series adjusts quickly to the data it is fed; this resulted in a sharpening of the classification through time as sensors with higher spatial resolution were introduced. The improvements in resolution from ASTER, Sentinel-1, and Sentinel-2 beginning in 2014 did not have any obvious effect on the accuracy measures, perhaps because the quality of the BULC products was already high.

4.5. Limitations in this study

Although this analysis was successful at creating a highly plausible LULC time series formed by synthesizing multiple sources of data, several limitations were apparent. The BULC classification's representation of Disturbed was limited by a confluence of factors between 1973 and 1979. These years were good for classifying Water and Forest, but quite poor for the Producer's Accuracy of the Disturbed class, which was low for several reasons. First, by chance there were no useable images in this region for 1974, making an update impossible in that year. Second, agriculture appears to have increased considerably in that period. Third, the expansion came at a time when there had been very little prior evidence of Disturbed land in the study area. Given these three obstacles, BULC was (temporarily) unable to produce a classification that scored a high Producer's Accuracy for Disturbed for part of 1975. With no images during that period, a large amount of land was disturbed was established without being concurrently observed. When a period of clear imagery returned, BULC needed to “catch up” to a better LULC map, which took several iterations for the new disturbed areas to be confirmed. Because there was very little of the that class before that point, errors in the newest areas played a heavy role in lowering the Producer's Accuracy values until enough new imagery arrived. For users interested in noting the precise timing of changes, however, disturbance dates should be interpreted with caution, in that the exact year of a change might be missed by one or even two years, depending on the rarity of new data. Within this study, that poor MSS-era performance illustrates the challenge of LULC mapping in low-data situations.

Users interested in more accurate timing of LULC changes in low-data conditions have several options within this analysis framework. First, it would be reasonable to assess the accuracy of BULC not at every incorporation of an Event, but on an annual basis as in many other time-series studies (e.g., Hawbaker et al., 2017; Song et al., 2014). Second and more interestingly, one could choose a second parameterization for BULC that is different for eras when data was known to be sparse. In this study, the timing noted for conversions to Disturbed was typically delayed by several Events as BULC waited for confirmation of changes. This was as intended, given that BULC was parameterized to not react too strongly to each new Event, in favor of stability in avoiding a classification series fraught with noisy, incorrect transitions. One could add the option of a specialized MSS-era parameterization that would be more trusting of each Event. This would improve the timing of disturbances but would, however, also have the effect of more errors of commission, as stray errors in Events might be immediately accepted by BULC; in general, the user could choose the level of acceptable noise. Separately, one could expand the functioning of the sequence detector to detect when a confirmed LULC change had first been seen in the series of Events. More generally, however, data rarity was less of a limitation in recent years and can be expected to lessen further as data becomes much more available as new instruments are launched, provided data access remains free. In processing future imagery from different platforms, it seems much more likely that the delay of a few Events might translate to a confirmation delay of a few weeks.

5. Conclusion

When Theodore Roosevelt and his party passed through this area a century ago it was, apparently, not so different from how the region appeared six decades later, at the dawn of the satellite age in 1972. The earliest Landsat images and associated BULC classifications, for 1972–3, show only one bit of estimated agriculture. By 1980, however, large-scale land clearing had already begun to change regional-scale LULC patterns. The intervening four decades, unlike the previous six since Roosevelt's journey, were times of remarkable change in the region: between 6 and 10% of the landscape appeared to be in active agriculture in 2016, a number that may seem small but had created obvious, visible, and long-lasting patterns on the landscape of the region.

The Bayesian Updating of Land Cover algorithm has several features that should be of interest in remote sensing's new age. Perhaps most importantly, as a sensor-independent algorithm it is able to benefit directly from the rich legacy of sensor-specific techniques developed over decades of scientific effort. In this study, simple CART models produced each Event; other users of BULC might choose to produce Events using more sophisticated techniques they deem useful, such as texture analyses, band transformations, and other techniques developed in the context of one or two sensors. By working with classified maps, BULC can rely on analysts to choose the appropriate legend, image quality, and classification strategy for a particular study in order to produce Events to a standard acceptable for the project. BULC's structure should be able to create time series that synthesize data from virtually any source. For tropical settings like that explored here, for example, it would be simple to gather more than the self-imposed limit of three images annually in recent years, and to more deeply employ radar data, which may be of substantial use where multiple optical sensors frequently return cloudy images. Using BULC could open time series analysis to rarely employed but high-quality satellite data, updating the time series for a pixel only when satisfactory evidence arrives from any sensor. In its automatic tracking of the estimated rise and fall of probabilities of these categories, BULC provides a clear mechanism for managing a vector of probabilities through time. In settings with, say, 5 or 10 classes, BULC can maintain an evidence-based estimate of probabilities at each time step, recording how more than one class's probability might increase in light of new data, while others decrease.

This study presented a use case for tracking deforestation over the entire satellite record, employing a very wide range of sensors from 1972 to 2016 in a cloudy region. Future applications could include either land or water monitoring over larger areas, integrating data from an arbitrary number of sensors of different types. As more and more remote sensing data from both the past and future are brought online, BULC can help to separate signal from noise, leveraging the unique

perspectives of each sensor to synthesize decades of views of Earth's surface.

Acknowledgements

This work was funded by an NSERC Undergraduate Student Research Award to Fortin and an NSERC Discovery Grant to Cardille. Three reviewers provided intriguing, incisive comments and questions about the work that greatly improved the manuscript. Stephen Stehman contributed invaluable time and effort to determine the proper statistical assessment during the revision process. Thanks to Jaaved and Max, for their many hours spent peering through clouds to label LULC points. Finally, thanks to the Google Earth Engine team, for their foresight to develop Earth Engine and their commitment to its success.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2019.111266.

References

Baccini, A., Goetz, S.J., Walker, W.S., Laporte, N.T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P.S.A., Dubayah, R., Friedl, M.A., 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nat. Clim. Chang. 2, 182.

Barlow, J., Lennox, G.D., Ferreira, J., Berenguer, E., Lees, A.C., Mac Nally, R., Thomson, J.R., de Barros Ferraz, S.F., Louzada, J., Oliveira, V.H.F., 2016. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. Nature 535, 144.

Bontemps, S., Defourny, P., Bogaert, E.V., Arino, O., Kalogirou, V., Perez, J.R., 2011. GLOBCOVER 2009-Products Description and Validation Report.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont, CA.

Brennan, R., Webster, T.L., 2006. Object-oriented land cover classification of lidar-derived surfaces. Can. J. Remote. Sens. 32, 162–172. https://doi.org/10.5589/m06-015.

Câmara, G., Valeriano, D.d.M., Soares, J.V., 2006. Metodologia para o Cálculo da Taxa Anual de Desmatamento na Amazônia Legal. pp. 24 (Sao Jose dos Campos).

Cardille, J.A., Foley, J.A., 2003. Agricultural land-use change in Brazilian Amazônia between 1980 and 1995: evidence from integrated satellite and census data. Remote Sens. Environ. 87, 551–562. https://doi.org/10.1016/j.rse.2002.09.001.

Cardille, J.A., Fortin, J.A., 2016. Bayesian updating of land-cover estimates in a data-rich environment. Remote Sens. Environ. 186, 234–249. https://doi.org/10.1016/j.rse.2016.08.021.

Cardille, J.A., Foley, J.A., Costa, M.H., 2002. Characterizing patterns of agricultural land use in Amazonia by merging satellite classifications and census data. Glob. Biogeochem. Cycles 16, 11–18. https://doi.org/10.1029/2000GB001386.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46. https://doi.org/10.1016/0034-4257(91)90048-B.

De Sy, V., Herold, M., Achard, F., Asner, G.P., Held, A., Kellndorfer, J., Verbesselt, J., 2012. Synergies of multiple remote sensing data sources for REDD+ monitoring. Curr. Opin. Environ. Sustain. 4, 696–706. https://doi.org/10.1016/j.cosust.2012.09.013.

DeFries, R., Achard, F., Brown, S., Herold, M., Murdiyarso, D., Schlamadinger, B., de Souza, C., 2007. Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. Environ. Sci. Pol. 10, 385–394. https://doi.org/10.1016/j.envsci.2007.01.010.

Earth Science and Remote Sensing Unit. NASA Johnson Space Center, 2018. Gateway to Astronaut Photography of Earth. https://eol.jsc.nasa.gov.

Fearnside, P.M., 2005. Deforestation in Brazilian Amazonia: history, rates, and consequences. Conserv. Biol. 19, 680–688.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4.

Fortier, J., Rogan, J., Woodcock, C.E., Runfola, D.M., 2011. Utilizing temporally invariant calibration sites to classify multiple dates and types of satellite imagery. Photogramm. Eng. Remote Sens. 77, 181–189.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sens. Environ. 122, 66–74. https://doi.org/10.1016/j.rse.2011.08.024.

Hansen, M.C., Stehman, S.V., Potapov, P.V., Loveland, T.R., Townshend, J.R.G., DeFries, R.S., Pittman, K.W., Arunarwati, B., Stolle, F., Steininger, M.K., Carroll, M., DiMiceli, C., 2008. Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. Proc. Natl. Acad. Sci. 105, 9439–9444. https://doi.org/10.1073/pnas.0804042105.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853. https://doi.org/10.1126/science.1244693.

Harris, N.L., Brown, S., Hagen, S.C., Saatchi, S.S., Petrova, S., Salas, W., Hansen, M.C., Potapov, P.V., Lotsch, A., 2012. Baseline map of carbon emissions from deforestation in tropical regions. Science 336, 1573–1576.

Hawbaker, T.J., Vanderhoof, M.K., Beal, Y.J., Takacs, J.D., Schmidt, G.L., Falgout, J.T., Williams, B., Fairaux, N.M., Caldwell, M.K., Picotte, J.J., Howard, S.M., Stitt, S., Dwyer, J.L., 2017. Mapping burned areas using dense time-series of Landsat data. Remote Sens. Environ. 198, 504–522. https://doi.org/10.1016/j.rse.2017.06.027.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., Vandriel, J.N., Wickham, J., 2007. Completion of the 2001 national land cover database for the conterminous United States. Photogramm. Eng. Remote Sens. 73, 337–341 (doi: citeulike-article-id:4035881).

Houghton, R.A., 1991. Tropical deforestation and atmospheric carbon dioxide. In: Tropical Forests and Climate. Springer, Dordrecht, pp. 99–118.

Houghton, R.A., 2012. Carbon emissions and the drivers of deforestation and forest degradation in the tropics. Curr. Opin. Environ. Sustain. 4, 597–603. https://doi.org/10.1016/j.cosust.2012.06.006.

Houghton, R.A., Skole, D.L., Nobre, C.A., Hackler, J.L., Lawrence, K.T., Chomentowski, W.H., 2000. Annual fluxes of carbon from deforestation and regrowth in the Brazilian Amazon. Nature 403, 301–304. https://doi.org/10.1038/35002062.

House, M.N., Wynne, R.H., 2018. Identifying forest impacted by development in the Commonwealth of Virginia through the use of Landsat and known change indicators. Remote Sens. 10. https://doi.org/10.3390/rs10010135.

Instituto Nacional de Pesquisas Espaciais (INPE), 2017. Annual Deforestation Rate in the Brazilian Legal Amazon (PRODES). http://www.obt.inpe.br/prodes/dashboard/prodes-rates.html.

Loveland, T.R., Belward, A.S., 1997. The IGBP-DIS global 1 km land cover data set, DISCover: first results. Int. J. Remote Sens. 18, 3289–3295. https://doi.org/10.1080/014311697217099.

Malingreau, J.P., Eva, H.D., de Miranda, E.E., 2012. Brazilian Amazon: a significant five year drop in deforestation rates but figures are on the rise again. Ambio 41, 309–314. https://doi.org/10.1007/s13280-011-0196-7.

Millard, C., 2005. River of Doubt: Theodore Roosevelt's Darkest Journey, 1st ed. Doubleday, New York.

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015.

Pearlman, J.S., Barry, P.S., Segal, C.C., Shepanski, J., Beiso, D., Carman, S.L., 2003. Hyperion, a space-based imaging spectrometer. IEEE Trans. Geosci. Remote Sens. 41, 1160–1173.

Powell, R.L., Matzke, N., de Souza, C., Clark, M., Numata, I., Hess, L.L., Roberts, D.A., 2004. Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. Remote Sens. Environ. 90, 221–234. https://doi.org/10.1016/j.rse.2003.12.007.

Roosevelt, T., 1914a. Col. Roosevelt's exploration of a tributary of the Madeira. Bull. Am. Geogr. Soc. 46, 512–519. https://doi.org/10.2307/200915.

Roosevelt, T., 1914b. Through the Brazilian Wilderness. C. Scribner's Sons, New York, NY.

Roosevelt Memorial Association Film Library, 1928. The River of Doubt. 2 Film Reels of 2 (799 Ft.). Roosevelt Film Library, United States.

Saatchi, S.S., Soares, J.V., Alves, D.S., 1997. Mapping deforestation and land use in Amazon rainforest by using SIR-C imagery. Remote Sens. Environ. 59, 191–202.

Savage, S., Lawrence, R., Squires, J., Holbrook, J., Olson, L., Braaten, J., Cohen, W., 2018. Shifts in forest structure in Northwest Montana from 1972 to 2015 using the Landsat archive from multispectral scanner to operational land imager. Forests 9, 157. https://doi.org/10.3390/f9040157.

Skole, D., Tucker, C., 1993. Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988. Science 260, 1905–1910. https://doi.org/10.1126/science.260.5116.1905.

Song, X.-P., Huang, C., Sexton, O.J., Channan, S., Townshend, R.J., 2014. Annual detection of forest cover loss using time series satellite measurements of percent tree cover. Remote Sens. 6. https://doi.org/10.3390/rs6098878.

Stehman, S.V., 2009. Sampling designs for accuracy assessment of land cover. Int. J. Remote Sens. 30, 5243–5272. https://doi.org/10.1080/01431160903131000.

Stehman, S.V., 2014. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. Int. J. Remote Sens. 35, 4923–4939. https://doi.org/10.1080/01431161.2014.930207.

Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morisette, J.T., Stehman, S.V., Woodcock, C.E., 2006. Global Land Cover Validation: Recommendations for Evaluation and Accuracy Assessment of Global Land Cover Maps. EUR 22156 EN - DG. Office for Official Publications of the European Communities, Luxembourg (48 pp.).

Thomlinson, J.R., 1999. Coordinating methodologies for scaling landcover classifications from site-specific to global: steps toward validating global map products. Remote Sens. Environ. 70, 16.

USDA FSA, 2018. NAIP Imagery. http://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/index.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., Driel, J.N.V., 1993. Completion of the 1990's national land cover data set for the conterminous United States. Photogramm. Eng. Remote. Sens. 67, 650–662.

Whitmore, T.C., Sayer, J., International Union for Conservation of Nature and Natural Resources. General Assembly, International Union for Conservation of Nature and Natural Resources. Commission on Ecology, IUCN Forest Conservation Programme,

1992. Tropical Deforestation and Species Extinction, 1st ed. Chapman & Hall, London; New York.

Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., Gobakken, T., 2012. Lidar sampling for large-area forest characterization: a review. Remote Sens. Environ. 121, 196–209. https://doi.org/10.1016/j.rse.2012.02.001.

Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., Hermosilla, T., 2018. Land cover 2.0. Int. J. Remote Sens. 39, 4254–4284. https://doi.org/10.1080/01431161.2018.1452075.

Xin, Q., Olofsson, P., Zhu, Z., Tan, B., Woodcock, C.E., 2013. Toward near real-time monitoring of forest disturbance by fusion of MODIS and Landsat data. Remote Sens. Environ. 135, 234–247. https://doi.org/10.1016/j.rse.2013.04.002.

Yin, H., Prishchepov, A.V., Kuemmerle, T., Bleyhl, B., Buchner, J., Radeloff, V.C., 2018. Mapping agricultural land abandonment from spatial and temporal segmentation of Landsat time series. Remote Sens. Environ. 210, 12–24. https://doi.org/10.1016/j.rse.2018.02.050.

Zhu, Z., 2017. Change detection using Landsat time series: a review of frequencies, preprocessing, algorithms, and applications. ISPRS J. Photogramm. Remote Sens. 130, 370–384. https://doi.org/10.1016/j.isprsjprs.2017.06.013.

Zhu, Z., Woodcock, C.E., 2014a. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: an algorithm designed specifically for monitoring land cover change. Remote Sens. Environ. 152, 217–234. https://doi.org/10.1016/j.rse.2014.06.012.

Zhu, Z., Woodcock, C.E., 2014b. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171. https://doi.org/10.1016/j.rse.2014.01.011.